Organisational learning from software project success

Claes Wohlin, Martin Höst and Björn Regnell

Abstract

This paper presents a method for using subjective factors to evaluate project success. The method is based on collection of subjective measures with respect to project characteristics and project success indicators. The paper introduces a new classification scheme for assessing software projects. The classification scheme is illustrated in a case study using data from 12 industrial projects. The results are positive and encouraging for future development of the approach.

1. Introduction

Information may be lost in software organisations since the human knowledge of software projects seldom is captured properly. Measurement of software projects and their success is a difficult undertaking. Many times people "know" things that are very hard to measure with objective measures of effort, defects and so forth. This could be exemplified with project competence or project management performance. This paper highlights how knowledge and experience of software developers, managers and customers may be captured and used. More specifically this paper illustrates how subjective measures may be used as a complement to, for example, effort and defect data.

A subjective measure may be defined as a measure that is solely dependent on the knowledge and expertise of the people involved. A potential advantage of a subjective measure is that it is fairly easy to collect through interviews or questionnaires, and no extensive measurement program has to be in place. The obvious disadvantage is that a subjective measure may be less exact and it may be hard to use the measures to draw conclusions. Instead of only focusing on the use of objective measures, we would like to propose using subjective measures as a complement to the collection of objective measures. The, perhaps, main use of subjective measures in software engineering has been in conjunction with effort and cost estimation, see for example [2, 6].

The following notion is used and an example is given below for illustration purposes. A subjective factor is a general term for an aspect we would like to study. The factor is either a project factor or success factor. A project factor may be divided into a number of project characteristics, and a success factor into a number of success indicators. Project characteristics provide a view of the status or quality of the project, and they can either be estimated prior to starting the project or during the execution of the project. A success indicator captures the outcome of the project, and it is hence measured after project completion. The project characteristics and success indicators are measured through variables. A subjective measure is then defined to measure the variables.

Example: An example of a project factor may be project management. This factor may include project characteristics such as the quality of the project plan and experience of the project manager. Examples of success indicators include timeliness of delivery and quality of the delivered software. The experience of the project manager may be studied through different variables, for example, number of times as project leader or through a survey among participants in previous projects. The first variable may be measured through calculating an absolute number, but it is also possible that it is judged that it is better to capture the experience on a five-point scale, since the differences between having been project leader, for example, 10 or 11 times are negligible.

A subjective evaluation method for using subjective measures in order to judge which project characteristics are most essential for project success has recently been developed [11].

The method provides input to project planning and control as well as risk management [5], where the method could provide valuable input to which projects are particularly difficult to turn into a success. It should not be viewed as a stand-alone method, but rather as a complement and support to project and risk management techniques used today. The method is summarised in Section 2.

The method is then extended to include a classification scheme for software projects. The classification is based on the rankings obtained using the subjective evaluation method. The extension is presented in more detail Section 3.

The extended method is illustrated and evaluated using data collected from 12 industrial software projects from one organisation, where 10 project characteristics were judged based on project reports and interviews with project participants. The characteristics are measured on a scale from 1 to 5. The success of the projects is measured through a project efficiency measure, which is defined as the development time divided by project effort. The efficiency is then mapped to a scale from 1 to 5. Nine projects are used to build the models needed to use the method. The remaining three projects are then used to illustrate how the method may be used when being applied in project planning and control. The case study is presented in Section 4.

In summary, the paper shows how subjective measures can be used successfully to plan and control software projects to increase the likelihood of project success. Some conclusions from the method and the case study are presented in Section 5.

2. Basic evaluation method

This section presents a summary of the evaluation method. This means that some details have intentionally been omitted. A full description of the method is available in [11].

The first step is to identify which subjective factors to evaluate and identify project characteristics and success indicators that are believed to be important. Suitable variables should be identified for both project characteristics and success indicators. It must be decided which project and success variables should be evaluated subjectively, i.e. measured on the subjective scales defined. It is usually better if the variables are measured quantitatively, and hence it is essential to determine what to measure quantitatively and what to evaluate subjectively. Success variables may include both internal (development-oriented) success variables, for example efficiency and maintainability, and external indicators (customer-oriented), for example timeliness and quality. Thus, it is important to decide what to evaluate when judging the success of a project. In other words, it should be determined which success variables to measure. The decision may be based on certain objectives, for example the timeliness of delivery is the key issue, or the objective is simply to capture project success in general. The project variables believed to influence the success variables should be identified and it should be determined how to measure them.

The scales for the subjective variables must be determined carefully. Different subjective rating schemes exist, for example, Likert and ordinal scales [4]. The meaning of the different values on the scale should be determined and these should preferably be defined so that we obtain a good spread between projects. Methods to define scales reliably are described in [9].

The following procedure is proposed in [11] to study the relationship between project characteristics and project success using subjective evaluation factors. The following analysis steps are then conducted for each success variable.

1. Correlation analysis (data screening)

We assume that all project and success measures have been formulated so that a higher value on a subjective variable is better. Hence, it is expected that all project variables should have a positive correlation with the success variables. If that is not the case, then there are either some underlying aspects that were not captured, or the scale does not represent what we think. To address issues like this, it is useful to screen the data, i.e. project variables that do not have a positive correlation with the success variable are removed from the analysis.

2. Principal component analysis I (project variables only)

   Different measures are often highly correlated because they measure almost the same thing. Different techniques exist to be able to extract a useful subset. One such technique is Principal Component Analysis (PCA) [7]. PCA groups correlated into a number of principal components, where each principal component accounts for the maximum possible amount of the variance for the variables being analysed. The number of principal components extracted may vary depending on the data set and criteria for extraction. Each variable is through the PCA given a loading for each principal component. The loading is a measure of the variables correlation with the resulting principal component. When applying PCA to non-interval variables, it is necessary to be a little cautious. This may, for example, include not accepting the results from the analysis without making sure that the results correspond to intuitive expectations. In this step, the project variables are analysed to study which project variables vary together and whether it is possible to identify certain commonalties among the project variables that vary together.

3. Principal component analysis II (project and success variables)

   A second PCA is conducted. This time the success variable is included in the analysis. The objective is to identify which project variables that vary together with the success variable. The loadings in the principal component containing the success variable are of particular interest in this analysis. A project variable is considered being in the same principal component as the success variable if the project variable has its highest loading in this component.

4. Ranking and correlation

   The projects are ranked twice. First, the projects are ranked based on the success variable, and secondly they are ranked using the project variables in the same principal component as the success variable. The latter ranking is based on the sum of the project variables with a loading higher than a certain threshold. Finally, the Spearman correlation is determined to evaluate the level of relation between the two rankings.

5. Agreement index

   The two rankings reflect different ways for evaluating a project. The first ranking models the project success based on one success variable. The second ranking also describes the projects based on project success, but this time based on the values of key project variables. The rankings are used to classify projects. If we, in the long run, want to learn how to predict project success based on project variables, the classifications based on the two rankings must show some degree of agreement. Agreement can be measured by an agreement index, often referred to as kappa statistic [1]. In software engineering, the kappa statistic has, for example, been applied to inter-rater agreement of process assessments [3]. A brief description of the kappa statistic is provided in [11].

   To be able to understand the degree of agreement, it is necessary to interpret the kappa statistic on a scale describing the degree of agreement. Several such scales exist, but there are only minor differences. Three scales are presented in [3]. Here the benchmark suggested by Altman [1] is used.

Table 1: A mapping between the kappa statistic and an interpretation [1].

Kappa statistics	<0.20	0.21-0.40	0.41-0.60	0.61-0.80	0.81-1.00
Strength of agreement	Poor	Fair	Moderate	Good	Very good

The agreement index is normally calculated for the agreement between, for example, different people using the same scale. This is not the situation here, since the two rankings are based on different variables and scales, and hence the classifications based on the rankings are based on different scales. This means that the agreement index may become lower than if the same scales were possible to use. All the same, the kappa statistic is used to compare the two classifications i.e. one from project variables and the other from success variables. The use of the kappa statistic under these circumstances is an area for further studies.

These five steps form the basic evaluation method for subjective factors. In the above steps, it is assumed that projects are classified as being successful or unsuccessful based on project characteristics and success indicators. This results in a 2 * 2 matrix for which the kappa statistic are calculated. One drawback with this approach is that borderline projects are easily misclassified as a potentially successful or unsuccessful project, although it turns out that it should have been classified the other way around. In other words, the exact border between successful and unsuccessful increases the misclassification. To address this problem, it is possible to include a middle class that basically says that the project could become either successful or unsuccessful. This additional class is only needed when classifying the projects using the project characteristics, since after having measured the success indicators it is assumed that it is known whether the project became a success or not. This extension to the basic method is described in more detail in the subsequent section.

3. Green, yellow and red projects

The classification scheme divides projects into green, yellow or red based on their project characteristics. Threshold values between the different classes may be chosen by each specific organisation using the method. The scheme is illustrated by classifying the projects based on project characteristics into three equally large groups, i.e. upper third (green), middle third (yellow) and lower third (red). The projects are after completion classified as either green (upper half) or red (lower half) based on the success indicators. The yellow class is not used for completed projects, since it is assumed that it is possible to determine whether the project was a success or not.

The idea behind having yellow projects is that it should be possible to classify projects as being somewhere in the middle, i.e. it is difficult based on the available information to predict that they will be either green or red after completion. The use of three classes allows us to indicate the uncertainty in classifying projects. Most classification schemes try to make exact classifications, and the success of the schemes is evaluated based on their ability to make exact classifications, see for example [8]. This is an ambitious and commendable approach, but it is also very difficult to succeed in terms of having low misclassification rates. The approach, suggested in this paper, tries to partly avoid this problem by accepting that there is an uncertainty.

The classification of projects may hence be summarised in a table with three rows and two columns as shown in Table 2.

Table 2: A classification table.

		Classification based on success indicators	
		Green	Red
Classification based on project characteristics	Green		
	Yellow		
	Red		

A set of projects may now be used to evaluate which project characteristics should be used for the classification. This is done using the basic evaluation method outlined in Section 2. The projects may then be placed in a table as the one in Table 2, and it is then possible to evaluate the correctness of the classification, since the success indicators are known. These projects are referred to as the fit projects, since they are used to build the classification and hence determine the borders between the different classes. New projects should then be classified using the project characteristics. The classification may then help in planning and conducting the project in order to achieve whatever success indicators are most important. Examples of usage are further discussed in the case study presented in Section 4.

The extension of the method to include a yellow class also means that the calculation and interpretation of the kappa statistic has to change. Previously, in the basic evaluation method, it was simple to calculate the agreement index based on that projects were simply classified as either green or red for both project characteristics and success indicators. This is no longer the case, since the scales are different. Two options exist:

- The first option is to disregard the projects that are classified as yellow using the project characteristics, and calculate the kappa statistics only based on the projects classified as either green or red for both project characteristics and success indicators.
- Another option is to view projects classified as yellow as always being correctly classified. The use of a yellow class has given managers on different levels a warning that a project is at risk of becoming unsuccessful. It should be noted that this option would in most cases result in a higher kappa statistic than the first option.

Independently, the classification table provides valuable input in terms of expectations of the projects given that the project characteristics are not changed. In particular, it provides an opportunity of trying to improve certain project characteristics based on the importance of the projects becoming a success or not. Further, the project characteristics should preferably be tracked and re-estimated regularly in the project. Of course, it would be beneficial if all projects were successful, and may be they all are successes, but there are always projects that are more successful than others are. In other words, the method presented here provides important input for informed decisions and prioritisation between different projects.

Next the complete method, including the extension of the classification table, is illustrated in a case study.

4. Case study

The data in the case study is from 12 software projects from one company consisting of many organisations. The company works with projects in the telecommunication domain. Data is collected for 10 project variables that were judged as being critical in improving the predictability of time to market. Only one success variable is judged as being important in relation to this objective, i.e. the lead-time. However, to make the lead-times comparable, they are normalised with the project effort before being judged on a five-point scale.

The 10 project variables are Problem Complexity, Competence, Requirements Stability, Staff Turnover, Geographical Distribution, Method and Tools, Time Pressure, Information Flow, Top Management Priority and Project Management. Once again, it may be appropriate to emphasise that the objective has been to formulate the subjective ordinal scales in accordance with the intuition. This means that a higher value for a variable is always assumed to be better. The complete scales can be found in [10]. The scale is formulated based on experience from the company where the data is collected. The same person has collected the data and the sources of information are final reports from the projects in combination with interviews with some key personnel in the projects.

The five steps in the basic method are first conducted for 9 of the 12 projects to create an experience base that can be used when assuming new projects should be planned. To be fair, nine projects are chosen by random to create the experience base. The number of projects in the case study is fairly limited, but at least sufficient to illustrate the use of the method. This also implies that the actual randomisation may affect the results in terms of screening and the principal component analyses. A more comprehensive case study from another company is presented in [11]. The steps results in the following:

1. Data screening

The data screening resulted in that seven of the ten project characteristics are positively correlated with the success variable. Three variables are hereafter removed from the analysis based on this. The three variables with negative correlation are Competence, Geographical Distribution, and Methods and Tools. The highest correlation (0.73) is between Requirements Stability and the success variable, and second highest is with Priority (0.70). It should be noted that the following two principal component analyses can only be carried out if the number of subjective variables is less than the number of projects. If this is not the case, some variable has to be removed based on other criteria.

2. PCA I (Project variables only)

Step 1 of the method leaves seven variables to include in the first PCA. The analysis results in three principal components, which are shown in Table 3. As a rule of thumb, based on experience, the focus is on loadings above 0.7. Based on this, it may be observed that the first principal component includes Time Pressure, Information Flow, Priority and Project Management. The second principal component includes primarily Complexity, although Staff Turnover is also fairly high, and finally the third principal component consists of Requirements Stability, but it could be noted that the loading for Staff Turnover is fairly high also for this principal component. This is one reason why it is suitable to choose a loading threshold of 0.7. That is to ensure that a specific variable is placed clearly in one principal component.

It is also beneficial if the different principal components are interpretable. The first principal component is most closely related to management in general. The second principal component is more difficult to explain, but the third principal component is simple since it only includes one variable. This analysis has been conducted to understand the project characteristics. The next step is to investigate their relationship with the success variable closer.

Table 3: Results from the first PCA with three principal components.

Project	PC 1	PC 2	PC 3
Complexity	0.065	0.920	-0.062
Requirements stability	0.110	-0.111	0.943
Staff turnover	-0.091	-0.638	0.580
Time pressure	0.923	-0.006	-0.109
Information flow	0.877	-0.239	0.257
Priority	0.861	0.261	0.351
Project management	0.811	0.313	-0.163

3. PCA II (Project variables and success variable)

The success variable is now included in the analysis, and the results are shown in Table 4. Several things may be observed from the analysis. First, it is noted that the success variable is placed together with Requirements Stability. Secondly, it may be noted that the first principal component is similar although the loadings have changed slightly. The second and third principal components have changed positions, and Staff Turnover is now more clearly placed together with Complexity. Given the analysis here it is now possible to rank the projects.

Table 4: Results from the PCA when including the success variable in the analysis.

Variables	PC 1	PC 2	PC 3
Complexity	0.052	0.032	0.917
Requirements stability	0.011	0.901	-0.216
Staff turnover	-0.130	0.474	-0.712
Time pressure	0.916	0.061	0.049
Information flow	0.852	0.307	-0.262
Priority	0.811	0.479	0.237
Project management	0.832	-0.124	0.298
Success variable	0.336	0.897	6.308E-5

4. Ranking and correlation

The projects are ranked based on the project characteristic(s) in the same principal component as the success variable given the threshold of 0.7. This means, in this particular case, that the nine projects are ranked based on the Requirements Stability and the success variable. The Spearman correlation between the two rankings becomes 0.78 (corrected for ties in ranks). Thus, it seems, as Requirements Stability may be a good indicator of potential success.

5. Agreement index with new classification table

Based on the two rankings and given that projects are divided into successful and unsuccessful as discussed previously, the classification table in Table 5 is obtained. This table indicates the quality of fit between the predictive model (project characteristics) and the actual outcome (success variable). Since the rankings are only based on one variable, there are several ties, which influence where to draw the borders between the classes. Letting projects belong to the better class solves this. For example, three projects are ranked as number three and they are all classified as green projects based on project characteristics. This results in five projects being classified as green based on project characteristics. A similar case occurs based on the success variable. The ranks are given within parenthesis for each project with the first rank denoting the rank based on project characteristics.

The fit, indicated by Table 5, is very good. It is noticeable that no project is placed in Green/Red and Red/Green respectively. The agreement index may now be calculated. In this particular case, the agreement index actually becomes 1.0 for both options described above, due to that there is no projects in the cells Green/Red and Red/Green. This is fairly unexpected, although very positive. The kappa statistic is not likely to be as good for all data sets, but all the same it is encouraging. It can also be seen that the new classification table has, in this particular case, resulted in that one project that otherwise would have been misclassified is placed in the yellow class (P4). The other yellow projects would have been classified correctly.

Table 5: A classification table for the nine projects, denoted P1-P9.

		Classification based on success indicators	
		Green	Red
Classification based on project characteristics	Green	P6(1,1), P5(2,4), P1(3,4), P3(3,3) and P8(3,1)	
	Yellow	P4(6,4)	P2(6,7) and P9(6,9)
	Red		P(9,8)

The five steps of the method have now been gone through to create a classification table that can be used when new projects are to be executed. The use of the table can be evaluated using the three projects not included in the previous analysis. The ranks are re-calculated and the placement of the new projects determines their class. Initially, the actual value on the success variable is unknown and hence the planning of the project has to be done based on the classification from the project characteristics. The results are summarised in Table 6.

Table 6: Results from using the classification table.

	P10	P11	P12
Class from project characteristics	Green	Red	Yellow
Class from success indicator	Red	Red	Green

From Table 6, it may be noted that one of the new projects (P10) was expected to become a success, but became unsuccessful. In this particular case, it was not enough with good requirements stability; other factors have obviously made the project less successful. P11 is predicted to become a red project and this is true. If project P11 was really crucial and important, management could (based on the result of this analysis) try to ensure that the requirements become more stable. There is also an opportunity to talk to the customer and explain the probable outcome based on that unstable requirements are expected. The method has for P11 provided management with important input to their decision process. Finally, P12 is likely to become either successful or unsuccessful; it is classified as yellow. Thus, management has information that means that other precautions may be taken. This may, for example, mean assigning one of their best project leaders or performing a very thorough risk assessment.

The main objective of the latter reasoning is to indicate how the methods for subjective factors may be used to support software projects.

5. Summary and conclusions

This paper has briefly presented a method for using subjective factors in project evaluation and planning. The method may be used during the early phases of a project. At this stage of the project, it is necessary to try to judge a set of project characteristics on an ordinal scale to be able to use the proposed method. The method is based on that subjective factors are investigated in a number of projects to form an experience base that may be used in new projects. It is also expected that the experience base be updated as new projects are conducted.

In particular, this paper introduces a new classification table where projects may be classified into three classes based on project characteristics and two classes based on project outcome. The main benefit with the classification table is that it is possible to indicate uncertainty in the classification rather than just dividing projects into successful and unsuccessful respectively.

The case study, illustrating the method, showed that it is possible to use subjective evaluations to help in project planning. The case study is however small so further studies are needed, which hopefully support the findings in this paper. Further work includes new case studies, extension to several success indicators and further improvements of the method.

6. References

[1] D. Altman, "Practical Statistics for Medical Research", Chapman-Hall, 1991.

[2] L. C. Briand, K. El Emam and F. Bomarius, "COBRA: A Hybrid Method for Software Cost Estimation, Benchmarking, and Risk Assessment", Proceedings IEEE International Conference on Software Engineering, pp. 390-399, 1998.

[3] K. El Emam, "Benchmarking Kappa: Interater Agreement in Software Process Assessment", Empirical Software Engineering, Vol. 4, pp. 113-133, 1999.

[4] N. Fenton and S. L. Pfleeger, "Software Metrics: A Rigorous & Practical Approach", International Thompson Computer Press, 1996.

[5] E. M. Hall, "Managing Risk: Methods for Software Systems Development", Addison-Wesley, 1998.

[6] M. Höst and C. Wohlin, "An Experimental Study of Individual Subjective Effort Estimations and Combinations of the Estimates", Proceedings IEEE International Conference on Software Engineering, pp. 332-339, 1998.

[7] S. K. Kachigan, "Statistical Analysis: An Interdisciplinary Introduction to Univariate and Multivariate Methods", Radius Press, 1986.

[8] T. M. Khoshgoftaar and E. B. Allen, "A Comparative Study of Ordering and Classification of Fault-Prone Modules", Empirical Software Engineering: An International Journal, Vol. 4, No. 2, pp. 159-186, 1999.

[9] A. von Mayrhauser, "Software Engineering: Methods and Management", Academic Press, 1990.

[10] C. Wohlin and M. Ahlgren, "Soft Factors and Their Impact on Time to Market", Software Quality Journal, No. 4, pp. 189-205, 1995.

[11] C. Wohlin and A. von Mayrhauser, "Assessing Project Success using Subjective Evaluation Factors", Technical report, Dept. of Communication Systems, Lund University, Sweden, 1999.